**MS in Artificial Intelligence**

**Course: Natural Language Processing**

**Credits:** 3.0

**Prerequisites:** Python and Programming, Linear Algebra, Probability and Statistics, Basic Calculus, and Deep Learning Introductory Courses

***Instructor*:** *Sai Praveen Kadiyala*

**Email**: [saipraveen.kadiyala@yu.edu](mailto:saipraveen.kadiyala@yu.edu)

***TA:*** *Kanchan Maurya*

**Email**: kmaurya@mail.yu.edu

**COURSE DESCRIPTION**

The course will be an introductory course in natural language processing – the study of human language from a computational approach-- using Python in applications.

**COURSE LEARNING OUTCOMES**

The course objectives will include the following:

* Foundational understanding of natural language processing approaches: rule-based and machine based
* Ability to distinguish across the various techniques and basic theory among the NLP approaches
* Exposure to practical applications of Natural Language Processing using Python

**REQUIRED TEXTS & SUGGESTED READINGS**

***Required readings will be provided throughout the semester, but here are the textbooks that are often referred to:***

*Textbooks*

* [Speech and Language Processing (3rd ed. draft)](https://web.stanford.edu/~jurafsky/slp3/). Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed.).
* Howard, J., & Gugger, S. (2020). Deep Learning for Coders with Fastai and PyTorch.
* Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python.
* Jacob Eisenstein. [Natural Language Processing](https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf)
* Michael A. Nielsen. [Neural Networks and Deep Learning](http://neuralnetworksanddeeplearning.com/)

*Software Requirements*

Below are the software instance options that you may choose to install locally to your computer to conduct the work outlined in this class:

* A code editor (e.g., [Notepad ++,](https://notepad-plus-plus.org/) [Visual Studio,](https://code.visualstudio.com/) etc.)
* Python 3.0 and an associated IDE (e.g., Jupyter)
* [Google Colab](https://colab.research.google.com/)
* Libraries: Pytorch, fast.ai, pandas, NLTK

**ASSIGNMENTS & GRADING**

*Use this area to describe your breakdown on grades. Below is an example:*

|  |  |
| --- | --- |
| **Assignment** | **Grade Weight** |
| Participation in Live Sessions, Quizzes and Stick checks | 10% |
| Teachback Assignment | 25% |
| Assignments | 40% |
| Final Project | 25% |

**Participation in Live Sessions and Readings for Discussion.** Participation in live sessions will be 10% of your overall grade. Missing more than one class could significantly impact your ability to pass. Thus, it is imperative that you make these classes. You will also be assessed on your participation during class discussions which could range from guest lectures to class demonstrations, to asks to share your screen and demonstrate an in-class group activity. In some cases, there may be readings required in advance of the session. Additionally, students will be called upon to present back their assignment for the week. Not being prepared will impact both your assignment and participation grade. Finally, stick checks are sometimes embedded in each module and could take the form of a quiz, case study or brief exercise

**Teachback Assignment.** A teach-back assignment requires students to select from a list of topics in the class and prepare a 15- to 20-minute recorded presentation on that topic in depth. Students will also review and provide feedback on a colleague’s project as a required component of the Teach-back assignment.

**Assignments.** In this class, participation and completion of assignments are critical. Assignments are problem sets typically using Python and the mode of submission will typically be a written summary and a code output. In some cases, the assignment shall be a preparation for a presentation on a selected topic in the class or preparation for the Final Project.

**Final Project.** As a class, we will split up into teams. Teams develop a project that incorporates the spectrum of content covered in the class. The project will provide students with the opportunity to solve a real-world problem using NLP techniques. We will have three components to the project: Literature Review, Dataset Creation, and Final Project Delivery.

The project will have several components:

* Peer Score
* Design and Planning
* Development of the Solution
* Written working paper
* Presentation

**Yeshiva Grading Scale**

***Table

Description automatically generated***

**KATZ SCHOOL CLASS ATTENDANCE POLICY**

Students are expected to attend all scheduled classes in their entirety. Students who fail to fulfill this requirement will receive an academic penalty appropriate for the course work missed.

Students may not miss 30% or more of their scheduled class. If a student misses 30% or more of a course during the semester, they will receive a final grade of “F.” This grade will be reflected on the student’s official university transcript.

For programs within clinical components students may not miss 20% or more of any course, clinical or not. At the Katz School, this pertains to only to students in the Speech Language Pathology program. If a student misses 20% or more of a course during the semester, they will receive a final grade of “F.” This grade will be reflected on the student’s official university transcript.

If the student is absent because of a disability which is documented with the Office of Disability Services at Yeshiva, falls ill or there are other extenuating circumstances, the student must inform the instructor in advance. The instructor may require appropriate documentation to make any exception to this policy.

Course Schedule

The standard week for all courses in the program begins on Monday and concludes the following Sunday. However, weekly modules will open every Friday morning to allow you to get a head start on the next module's work. Please check the due dates in the course schedule for weekly assignments.

The schedule below includes a tailored set of materials for the topic chosen by the student.

| **Week** | **Title** | **Topic(s)** | **Readings and Assignments** |
| --- | --- | --- | --- |
| 1 | Week 1: Introduction to Natural Language Processing (NLP) | * Overview of NLP: history, applications, and challenges. * Basic linguistic concepts: morphology, syntax, and semantics. * Introduction to Python, Fast.AI, and PyTorch for NLP. * Review of NLP Tasks | * Module 1 * Assignment 1 Assigned * Teachback Assignment Assigned |
| 2 | Week 2: Text Preprocessing and Language Modeling | * Mathematics and Machine Learning Review * Text normalization: stemming, lemmatization, and stop word removal. * N-grams and language modeling. * Hidden Markov Models (HMM) for language generation. | * Module 2 * Assignment 1 Due * Assignment 2 Assigned |
| 3 | Week 3: Part-of-Speech Tagging and Named Entity Recognition | * Part-of-speech tagging: rule-based and statistical approaches. * Hidden Markov Models (HMM) for POS tagging. * Named Entity Recognition (NER): methods and evaluation. | * Module 3 * Assignment 2 Due * Final Project Assigned * Assignment 3 Assigned |
| 4 | Week 4: Syntax and Parsing | * Context-free grammars and parsing techniques. * Dependency parsing: transition-based and graph-based approaches | * Module 4 * Assignment 3 Due * Assignment 4 Assigned |
| 5 | Week 5: Sentiment Analysis and Opinion Mining | * Sentiment analysis: methods and applications. * Opinion mining: aspect-based sentiment analysis and opinion summarization. | * Module 5 * Assignment 4 Due * Assignment 5 Assigned |
| 6 | Week 6: Word Embeddings and Neural Language Models | * Deep learning concepts: * Recurrent Neural Network (intro) * Feed-forward network or Multi-Layer Perceptron (MLP) * Cross-entropy loss function and gradient descent optimization | * Module 6 * Assignment 5 Due * Assignment 6 Assigned |
| 7 | Week 7: Sequence-to-Sequence Models and Machine Translation | * Sequence-to-sequence models: encoder-decoder architectures. * Neural machine translation: attention mechanisms. | * Module 7 * Assignment 6 Due * Assignment 7 Assigned |
| 8 | Week 8: Question Answering and Dialog Systems | * Question answering systems: retrieval-based and generative models. * Dialog systems: rule-based and data-driven approaches. | * Module 8 * Assignment 7 Due * Assignment 8 Assigned * Teachback Assignment Due |
| 9 | Week 9: Neural Network Architectures for NLP | * Stacked and Bi-Directional RNN * Managing context in RNN: LSTM and GRU * Word order and lexical gap * Encoder-Decoder/Seq2Seq Model * BEAM search | * Module 9 * Assignment 8 Due * Assignment 9 Assigned |
| 10 | Week 10: Unsupervised and Semi-Supervised Learning for NLP | * Constituency * Narrow vs. Wide convolution | * Module 10 * Assignment 9 Due * Assignment 10 Assigned |
| 11 | Week 11: Advanced Topics in NLP | * Neural machine translation and multilingual NLP. * Text summarization: extractive and abstractive approaches. * Conversational AI and chatbots. * Large Language Models (LLM) * Natural Language Generation (NLG) and text-to-speech synthesis. * Ethical considerations in NLP: bias, privacy, and fairness. * Current trends and future directions in NLP research. | * Module 11 * Assignment 10 Due * No assignment for the following week |
| 12 | Final Project Presentations | * Final Project Presentation | * Final Project Due |

**ONLINE AND/OR BLENDED LEARNING POLICIES**

**Online Learning Engagement Policy**

A successful online class only happens when there is an active community. Students are required to

attend both the weekly live synchronous sessions and participate in other community building activities such as the discussion boards.

**Netiquette**

Netiquette is a set of rules for behaving properly in an online course. Often the anonymity of online courses can cause a lapse in judgement when learners are excited or passionate about a subject. This can lead to statements that could be demeaned as offensive. You are all adults and are treated as such. However, it is still important to talk about these issues. The following bullet points cover some basics communicating in an online course:

* Be sensitive to the fact that there will be people with different cultural and linguistic backgrounds, as well as different political and religious beliefs.
* Use good taste when composing your responses in Discussion Forums. Swearing and profanity is also part of being sensitive to your classmates and should be avoided.
* Don’t use all capital letters when composing your responses as this is considered “shouting” on the Internet and is regarded as impolite or aggressive.
* Be respectful of your others’ views and opinions. Avoid “flaming” (publicly attacking or insulting) them as this can cause hurt feelings and decrease the chances of getting all different types of points of view.
* Be careful when using acronyms. If you use an acronym it is best to spell out its meaning first, then put the acronym in parentheses afterward, for example: Frequently Asked Questions (FAQs). After that you can use the acronym freely throughout your message.
* Use good grammar and spelling (avoid using text messaging shortcuts).
* If you aren’t sure what someone meant, consider asking for clarification.
* Remember that your peers are not required to respond to your specific post, so don’t be offended if your question goes unanswered.

**UNIVERSITY POLICIES AND RESOURCES**

**KATZ SCHOOL POLICIES**

[Click here](https://www.yu.edu/registrar/grad-catalog) and select “Katz School” Policies.

**ACCESSIBILITY AND ACCOMMODATIONS**

The Office of Disability Services collaborates with students, faculty and staff to provide reasonable accommodations and services to students with disabilities. Students with disabilities who are enrolled in this course and who will be requesting documented disability-related accommodations should make an appointment with the Office of Disability Services by calling (646) 592-4132 or emailing [rkohn1@yu.edu](mailto:rkohn1@yu.edu), during the first week of class. Once you have been approved for accommodations, please submit your accommodation letter to ensure the successful implementation of those accommodations. For more information, please visit: <https://www.yu.edu/Student-Life/Resources-and-Services/Disability-Services>

**ACADEMIC INTEGRITY**

The submission by a student of any examination, course assignment, or degree requirement is assumed to guarantee that the thoughts and expressions therein not expressly credited to another are literally the student’s own. Evidence to the contrary will result in appropriate penalties.

Academic integrity is a set of responsibilities and standards to facilitate high academic quality and rigor with the purpose of clarifying expectations and student conduct. The submission by a student of any coursework, or degree requirement is assumed to guarantee that the thoughts and expressions therein not expressly credited to another are literally the student’s own. Examples of violations on academic integrity are, but not limited to:

* Cheating
* Plagiarism
* Dishonesty
* Assisting or attempting to assist another student in an act of academic dishonesty
* Providing papers, essays, research, or other work to aid another student in Intentional Misrepresentation
* Engaging in unauthorized cooperation with other individuals in completing assignments or examinations
* Submitting the same assignment, in part or whole, in more than one course, whether at YU or another institution, without prior written approval from both faculty members.

For more information, visit <http://yu.edu/registrar/grad-catalog/>

**YU Refund Policy**

You should be aware of the university's refund policy. Please review this information: <https://www.yu.edu/osf/undergraduate-accounts/withdrawal>.

**Academic Calendar**

You should review the academic calendars, including add/drop dates. Please review this information: <https://www.yu.edu/registrar/grad-calendar>.

**STUDENT SUPPORT SERVICES**

Katz School offers academic support through the Learning Hub. This support service includes writing, academic integrity (APA format), English as a Second Language, and general academic tutoring. For more information, please contact [katz@yu.edu](mailto:katz@yu.edu).

If you need any additional help, please visit Student Support Services at <http://yu.edu/academics/services/>